

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Diplomová práce

Nástroj pro automatickou identifikaci KIR alel

Místo této strany bude
zadání práce.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 12. února 2020

Kateřina Kratochvílová

Poděkování

Ráda bych poděkovala Ing. Lucii Houdové, Ph.D. za cenné rady, věcné připomínky, trpělivost a ochotu, kterou mi v průběhu zpracování této práce věnovala.

Abstract

The text of the abstract (in English). It contains the English translation of the thesis title and a short description of the thesis.

Abstrakt

Text abstraktu (česky). Obsahuje krátkou anotaci (cca 10 řádek) v češtině. Budete ji potřebovat i při vyplňování údajů o bakalářské práci ve STAGu. Český i anglický abstrakt by měly být na stejné stránce a měly by si obsahem co možná nejvíce odpovídat (samozřejmě není možný doslovný překlad!).

Obsah

1	Úvod	7
2	Geny	8
2.1	HLA a non-HLA geny	8
2.1.1	Jak vypadá genom	8
2.2	Natural Killer	9
2.3	Jak funguje HLA	10
2.4	Jak funguje non-HLA	10
2.5	Bordel pro první kapitulu	10
2.6	Sekvence DNA	12
3	Sekvenační metody získávání DNA dat	13
3.1	Porovnání vhodného dárce	13
3.2	Sanger sequencing	14
3.3	NGS next-generation sequencing	14
4	Analyza dostupných bioinformatických nástrojů pro zpracování NGS dat	15
4.1	ART	15
	Literatura	16

1 Úvod

2 Geny

V každé buňce lidského organismu, konkrétně v buněčném jádře, je možné nálezt 46 chromozomů. Jeden chromozom představuje stočenou dlouhou molekulu DNA (Deoxyribonuklenovou kyselinu). Všechny 46 chromozomů obsahuje okolo 100 000 genů. Drobný segment DNA, který řídí buněčnou funkci je právě gen. Konkrétní forma genu je alela. [8]

2.1 HLA a non-HLA geny

Human leucocyte antigen(HLA) je genetický systém, který je primárně zodpovědný za rozeznávání vlastního od cizorodého. Tento systém je složen právě z jednotlivých HLA genů rozpoznávajících antigeny (cizorodé částice). Pokud HLA gen přijde do styku s antigenem je antigen zničen. Non-HLA geny jsou geny které se nepodílejí na základní funkci HLA systému. základní rozdíl mezi HLA a non-HLA a kir

Non-HLA geny jsou geny které se nepodílejí na základní funkci HLA systému. Z III třídy jsou to všechny, z II žádný a z I je to směs. Zjednodušeně můžeme říci, že geny které nejsou HLA jsou non-HLA. Tyto geny souvisejí též s funkcí imunitního systému, ne však vylučně s funkcí HLA.

2.1.1 Jak vypadá genom

Genová oblast HLA komplexu, se nalézá na krátkém raménku 6. chromozomu (6p21.31), zaujímá úsek dlouhý 3600 kb (3,6cM), tedy přibližně jednu tisícinu genomu. Obsahuje 224 genů; 128 funkčních genů a 96 pseudogenů a patří k regionům s nejvyšší genovou hustotou.

Uprostřed HLA oblasti se nachází úsek o velikosti 1 Mb, ve kterém bylo identifikováno na 70 genů, které se funkčně ani strukturně nepodobají HLA molekulám. Navzdory této skutečnosti se vžil označení geny III. třídy, přičemž některé geny původně zařazené do této třídy jsou nověji označovány jako geny IV. třídy (viz. výše).

HLA-6.Chromozom a KIR 19.chromozom udíží se segregují nezávisle a HLA shodní dárce s příjemce mají obvykle různé složení KIR genů (Fryčová)

2.2 Natural Killer

Velká buňka imunitního systému, nepotřebuje antigen aby začala zabíjet. -nespecifická imunita - vrozená, neadaptivní - veškeré potřebné informace zapsaná v DNA. Odpovídá při každém setkání s antigenem stejně - nemá paměť -> tedy si to pročiřečí KIR jsou receptory na povrchu NK buněk, NK zabíjejí na základě interakce mezi KIR receptorem a HLA molekulou na povrchu buňky

NK buňky mají schopnost identifikovat buňky vlastního MHC systému (HLA I.třídy) které jsou normálně exprimovány prakticky na všech buňkách v těle. Nádorové a některé virem napadené buňky potlačují expresi HLA I. třídy a tím se brání napadení cytotoxickými T lymfocyty. Molekuly HLA I. třídy rozpoznávají NK buňky pomocí pozitivních a negativních receptorů, které mohou inhibovat nebo naopak aktivovat NK buňky k „zabíjení“

V užším slova smyslu se jako ligand označuje signální molekula, která se váže na vazebné místo cílového proteinu. Ligand, který je schopný po navázání na receptor vyvolat fyziologickou odpověď, se nazývá agonista, ten, který je schopen se vázat, ale odpověď nespouští, je antagonist

Zjednodušeně: NK buňky neustále systematicky zjišťují přítomnost či absenci příslušný HLA ligand pro své KIR receptory. Pokud je HLA molekula přítomna, pak dojde k vazbě KIR-ligand HLA a protože za normální okolností převládají inhibiční KIR nad aktivačními, tak nedojde ke spuštění cytotoxické reakce NK buněk. Jestliže receptory KIR nenalzenou příslušný ligand HLA (vlastní molekulu HLA) aktivační KIR receptory převládnu nad inhibičními a je spuštěna náležitá cytotoxická kaskáda. lymfocyty bílá krvinka je leukocyt - typ bílé krvinky - T a B lymfocyty - specifická imunita - NK buňky nespecifická imunita - vznikají v z lymfatických kmenových buněk v kostní dřeni Aha takže lymfatické řečiště je více propustné proto to co nejde do cév jde sem pak se to odfiltruje a pak se to vrací do krevního řečiště.

KIR jsou na povrchu NK buněk a kde jsou teda NK buňky? NK je v podstatě lymfocyt a to je typ bílé krvinky. jo a nebudou teda spíš v lymfatické uzlině? leukocyty 1. granulocyty - neutrofilní, bazofilní a eozinofilní 2. agranulocyty - lymfocyty a monocyty

neutrofilní granulocyty jsou schopny vycestovat z kapilár do místa zánětu přeměněné monocyty přítomné v játrech v tělních dutinách (hrudní, břišní), ve slezině vy lymfatických uzlinách a kostní dřeni

2.3 Jak funguje HLA

2.4 Jak funguje non-HLA

2.5 Bordel pro první kapitolu

Takže to vypadá že nejdřív se najde shoda HLA a pak se ještě dodělává KIR shoda. Proč KIR? protože roste počet důkazů vlivu genů KIR že mají vliv na výsledky transplantace při leukemii HLA je na 6. chromozomu KIR je 19 chromozomu. tudíž se segregují nezávisle a HLA shodní dárce s příjemcem mají obvykle různé složení KIR genů Nesmírná variabilita alel tohoto systému ztěžuje úspěšnost allogeních transplantací.

HLA jen zkopírováno a je ta i hezkej obrázek z Genová oblast HLA komplexu, se nalézá na krátkém raménku 6. chromozomu (6p21.31), zaujímá úsek dlouhý 3600 kb (3,6cM), tedy přibližně jednu tisícinu genomu. Obsahuje 224 genů; 128 funkčních genů a 96 pseudogenů a patří k regionům s nejvyšší genovou hustotou.

Uprostřed HLA oblasti se nachází úsek o velikosti 1 Mb, ve kterém bylo identifikováno na 70 genů, které se funkčně ani strukturně nepodobají HLA molekulám. Navzdory této skutečnosti se vřilo označení geny III. třídy, přičemž některé geny původně zařazené do této třídy jsou nověji označovány jako geny IV. třídy (viz. výše).

HLA nomenklatura HLA nomenklatura - zase jen skopírováno Vysoký stupeň polymorfismu HLA systému zohledňují platné zásady pro označování HLA alel dané Světovou zdravotnickou organizací WHO (WHO nomenklatura). Princip je jednoduchý: Každá alela je definována písemným označením lokusu následovaným hvězdičkou (HLA-DRB1*), a poté kombinací 4 číslic (*0401), přičemž první dvojčíslí určuje sérologickou specifitu dané alely, druhé pak označuje alelu na základě její aminokyselinové sekvence. Případné páté číslo charakterizuje tzv. "tichou" variantu alely, tzn. záměnu nukleotidů bez změny aminokyselinové sekvence.

Dědičnost HLA geny jsou děděny autozomálně kodominantně a vykazují mendelistický typ dědičnosti. Počet rekombinací v HLA systému je řídký, vyskytuje se přibližně v 1 případě a častěji u žen. Celá oblast od HLA-F až po HLA-DP se přenáší z rodičů na potomstvo jako haplotyp. V rámci rodiny se mohou vyskytnout teoreticky 4 různé kombinace rodičovských haplotypů, takže sourozenci mohou být navzájem buď HLA identičtí, haploidentičtí (mají jeden haplotyp, v druhém se liší), anebo rozdílní. Rodiče jsou vůči svým dětem vždy haploidentičtí [5]. Z genetického hlediska

významný fenomén představuje existence vazebné nerovnováhy (linkage disequilibrium) v rámci HLA. Mnoho HLA genů se nalézá v tak těsné blízkosti, že se přenášejí z rodičů na potomky téměř vždy společně. V důsledku této skutečnosti se v populaci vyskytují některé kombinace alel různých genů častěji, než by se očekávalo. Vazebná nerovnováha je významným faktorem v asociaci HLA antigenů s chorobami, protože mnohá onemocnění se v jejím důsledku váží s více antigeny.

Non-HLA geny Non-HLA geny jsou geny které se nepodílejí na základní funkci HLA systému. Z III třídy jsou to všechny, z II žádný a z I je to směr. Zjednodušeně můžeme říci, že geny které nejsou HLA jsou non-HLA. Tyto geny souvisejí též s funkcí imunitního systému, ne však výlučně s funkcí HLA.

lymfocyty bílá krvinka je leukocyt - typ bílé krvinky - T a B lymfocyty - specifická imunita - NK buňky nespecifická imunita - vznikají v z lymfatických kmenových buněk v kostní dřeni Aha takže lymfatické řečiště je více propustné proto to co nejde do cév jde sem pak se to odfiltruje a pak se to vrací do krevního řečiště.

KIR jsou na povrchu NK buněk a kde jsou teda NK buňky? NK je v podstatě lymfocyt a to je typ bílé krvinky. jo a nebudou teda spíš v lymfatické uzlině? leukocyty 1. granulocyty - neutrofilní, bazofilní a eozinofilní 2. agranulocyty - lymfocyty a monocyty

neutrofilní granulocyty jsou schopny vycestovat z kapilár do místa zánětu přeměněné monocyty přítomné v játrech v tělních dutinách (hrudní, břišní), ve slezině vy lymfatických uzlinách a kostní dřeni

KIR KIR jsou teda jak na HLA tak na non-HLA? Je to součástí genu - řadí se do přirozené (nespecifické) imunity narozdíl od B-buněk a T-buněk. - NK buňky představují 10-15% lymfocitů v periferní krvi - jsou to buňky které reagují rychle a efektivně likvidují především nádorové buňky a buňky infokované virem

NK nemají antigenné specifické receptory, jak rozeznávají abnormální buňky? NK buňky identifikují molekuly vlastního MHC systému

jmenovitě HLA I. třídy, které jsou normálně exprimovány prakticky na všech buňkách v těle. Nádorové a některé virem napadené buňky potlačují expresi HLA I. třídy a tím se brání napadení cytotoxickými T lymfocyty (Restifo, 1993). Snížená exprese HLA I. třídy činí abnormální buňky citlivé k cytotoxicitě NK buněk (Karre, 1986). Molekuly HLA I. třídy rozpoznávají NK buňky pomocí pozitivních a negativních receptorů, které mohou inhibovat nebo naopak aktivovat NK buňky k „zabíjení“

Stručně lze shrnout, že NK buňky s potenciálem iniciovat cytotoxickou aloreakci používají KIR receptory jako inhibiční, směrem k „vlastním“, zdra-

vým buňkám. Pokud však příslušný vlastní ligand HLA na cílové buňce chybí, pak dochází k iniciaci cytotoxické reakce. Proces interakce KIR/HLA a mechanismus regulace cytotoxicity NK buněk se jako

receptory imunoglobulinové (protilátka - protein, který je schopen jako součást imunitního systému identifikovat a zneškodnit cizí objekty - bakterie a viry) v těle. Protilátky jsou nositeli humorální imunity. Jsou to krevní bílkoviny vznikající v mízní tkáni. povahy nacházejících se na povrchu Natural killers buněk a některých T-buněk (Variabilita v sekvenci).

KIR3D - prej tři skupiny ale to fakt divně popsany (českej článek) něco s imunoglobulinovými doménami KIR2D

funkce KIR -

these genes are encoded on chromosome 19. NK zabíjejí na základně interakce mezi KIR receptorem a HLA molekulou na povrchu buněk. Mohou mít různé podoby.

HLA i KIR jsou na různých chromozomech proto se segregují nezávisle a HLA schodni darci mají obvykle různé složení KIR genů

Struktura nukleových kyselin

jen skopírované z Nukleové kyseliny (polynukleotidy) jsou tvořeny dlouhými řetězci (mono)nukleotidů, vzájemně spojených fosfodiesterovými vazbami. Řadíme je k tzv.heteropolymérům, neboť jsou sestaveny z různých typů základních jednotek. Tato skutečnost je podstatná pro uchovávání a předávání informace, což je základní funkce nukleových kyselin v organismu. Homopolyméry (např. glykogen) obsahují pouze jeden typ monoméru (v našem případě glukózu), a tak nemohou plnit informační funkci.

2.6 Sekvence DNA

Je posloupnost písmen představující primární strukturu reálné nebo hypotetické molekuly či vlákna DNA, které má kapacitu nést informaci.

Používaná písmena A, C, G a T reprezentují čtyři nukleotidy ve vláknu DNA – adenin, cytosin, guanin a thymin, lišící se typem báze kovalentně vázané k fosfátové páteři. Posloupnost libovolného množství nukleotidů většího než čtyři lze nazývat sekvencí. Obvykle se sekvence vypisuje bez mezer, např. AAAGTCTGAC, ve směru 5 -> 3. Vzhledem k biologickým funkcím, které mohou záviset na kontextu, sekvence buďto mají anebo nemají smysl a jsou tedy kódující nebo nekódující DNA. Typem nekódující sekvence DNA je také tzv. „junk DNA“.

TO je z wiki bacha na to.

3 Sekvenační metody získávání DNA dat

Někdy se sekvenují pouze jisté části genomu které mají pro výzkumníka v daném okamžiku význam.

Sekvenování DNA je souhrnný termín pro biochemické metody, jímž se zjišťuje pořadí nukleových bází (A, C, G, T) v sekvenci DNA. Tyto sekvence jsou součástí dědičné informace v jádru. Adenin s thyminem a cytosin s guaninem.

zjišťování primární struktury nukleových kyselin (sekvenování)

Užitečné nejen ve výzkumu ale i v diagnostice nemocí či forenzní medicíně.

3.1 Porovnání vhodného dárce

V případě nepříbuzenských transplantací se vybírají potenciální dárce, kteří nemají s daným pacientem žádný děděný haplotyp. Snahou je najít takového dárce, který má shodné, přestože děděné od jiných rodičů, HLA antigeny. Informace o tom, jak jsou alely haplotypicky uspořádány obvykle chybí, proto je vždy nutná typizace maximálním rozlišením ve více HLA lokusech. Zjišťovaný minimální rozsah HLA shody se v jednotlivých transplantačních centrech liší. V současné době je u nepříbuzného páru požadována typizace vysokým rozlišením v lokusech HLA – A, B, C, DR a DQ (<http://www.efiweb.eu/efi-committees/standards-committee.html>). Pokud pacient a dárce mají stejné alely na všech těchto lokusech, hovoříme o shodě 10/10. Při jedné neshodě se jedná o shodu 9/10, při dvou o shodu 8/10. K typizaci se nejčastěji používá PCR – SSP (PCR se sekvenačně specifickými primery) či SBT (sequence based typing) technika, v posledních letech se stává zlatým standardem přímá sekvenace (SBT) HLA genu. fríčová U HLA - A, B, C, DR a DQ požaduje se typizace v těchto lokusech. Pokud má dárce shodu ve všech lokusech hovoříme o shodě 10/10 při jedné neshodě je to 9/10.

U Kir jsou 2 hlavní typy haplotyp A a B, které jsou definovány typem a počtem specifických KIR genů. Neexistuje žádné jednoduché univerzální kritérium definující a odlišující tyto haplotypy. Sekvenační metody s elší především rychlostí a cenou.

3.2 Sanger sequencing

K sekvenaci se používá gelová elektroforéza použitelná k sekvenování krátké sekvence jednovláknové DNA. využívá biologického procesu replikace DNA. Vybraná sekvence se vloží do reakční směsi s radioaktivně označeným primerem.

3.3 NGS next-generation sequencing

Je rychlé a relativně nenáročné zpracování jednotlivých vzorků. Tisíce až miliony sekvencí mohou být produkovány během jednoho sekvenčního procesu. K popularitě této metody nepomohla i komerční dostupnost stolních sekvenátorů.

4 Analyza dostupných bioinformatických nástrojů pro zpracování NGS dat

4.1 ART

ART je set simulačních toolů, kteří generují next generation sequence ready.
[2]

tak jsem stáhla normálně nejnovější verzi z niehs.nih.gov a podle instrukcí
co byli v souboru INSTAL dala

musí se brát v potaz že z toho generátoru nikdy nebudou data taková
jako reálná.. realná budou horší

na stránkách [elixir-europe](http://elixir-europe.org) což je oragnizace co má dávat dohromady
všechny vědecký veci a bla bla.

Tak tam je přímo Bowtie [4]

šla jsem přes docker `docker image ls` - zobrazí vsechny image pak `docker
run` a ID image `sudo docker run -i -t 3c2b9a287f82 /bin/bash` `sudo docker
ps -a`

Tak jsem nakonec žádněj docker nepotřebovala a stáhla jsem to tady po
kliknutí na bowtie binary release.

Literatura

- [1] FRYČOVÁ, M. Lze u pacientů s AML indikovaných k nepříbuzenské transplantaci provádět v klinické praxi výběr nepříbuzných dárců na základě KIR genotypů, 2016.
- [2] HUANG, W. et al. ART: a next-generation sequencing read simulator. 2012. Dostupné z: <https://academic.oup.com/bioinformatics/article/28/4/593/213322>.
- [3] J, R. et al. *Nomenclature* [online]. Nucleic Acids Research, 2015. [cit. 2019/10/1]. 43:D423-431. Dostupné z: <http://hla.alleles.org/misc/citing.html>.
- [4] LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. 2009. Dostupné z: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-3-r25>.
- [5] MUDR. PAVEL JINDRA, P. D. *Imunopatologické a imunogenetické aspekty transplantací krvetvorných buněk a solidních orgánů*. PhD thesis, Universita Karlova v Praze, 2011.
- [6] ROBINSON, J. et al. The IMGT/HLA Database. 2013. Dostupné z: <https://www.ebi.ac.uk/ipd/index.html>.
- [7] ROBINSON, J. et al. IPD—the Immuno Polymorphism Database. 2013. Dostupné z: <https://www.ebi.ac.uk/ipd/index.html>.
- [8] SMITH, D. T. *Encyklopedie lidského těla*. 2005. ISBN 80-7321-156-4.